Physiological mechanisms of masking

Bertrand Delgutte

Research Laboratory of Electronics,
Massachusetts Institute of Technology, Cambridge, Massachusetts,
and Eaton–Peabody Laboratory,
Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.

Introduction

The idea that psychophysical masking patterns produced by acoustic stimuli are closely related to their patterns of excitation along the cochlea (Wegel and Lane, 1924; Fletcher, 1940; Zwicker, 1970), has been useful in hearing theory. This correspondence is based on the view that masking is due to the spread of the excitation produced by the masker to the place of the tone signal along the cochlea. Studies of two-tone suppression in auditory-nerve fibers (Sachs and Kiang, 1968) challenged this classical view because masking might be due to the suppression of the neural responses to the signal by the masker, even if the masker does not excite neurons tuned to the signal frequency (Javel et al., 1983; Pickles, 1984; Sinex and Havey, 1986)\(^1\). This possibility that suppression contributes to masking when the signal and the masker are simultaneously presented led Houtgast (1974) to propose that nonsimultaneous masking techniques (such as pulsation thresholds) might better reflect the pattern of excitation produced by the masker than simultaneous masking. In this paper, we compare masked thresholds of auditory-nerve fibers obtained by simultaneous and nonsimultaneous techniques in order to separate the contributions of two-tone rate suppression and spread of excitation to tone-on-tone masking. The results show that physiological masking is both excitatory and suppressive, with the relative importance of the two mechanisms being dependent on the masker level and the frequency separation between signal and masker.

Figure 1 illustrates how spread of excitation and suppression might mask the response of an auditory-nerve fiber to a tone signal. Each panel shows discharge rate as a function of signal level both in the presence and in the absence of a fixed masker. In Fig. 1A, the masker is excitatory because it produces an increase in discharge rate over spontaneous. Two-tone rate suppression does not occur because the rate in response to the signal plus the

\(^1\text{Either of two response measures might be suppressed, the average rate of discharge, or the synchrony of discharges to the signal frequency. This paper concerns the role of two-tone RATE suppression in masking.}

204
masker is never lower than the rate for the signal alone. The physiological masked threshold is the level at which the rate produced by the signal plus the masker exceeds the rate for the masker alone by a certain probabilistic criterion. Masking is said to occur because the masked threshold is higher than the threshold in quiet, which is the level for which the response to the signal alone exceeds spontaneous rate by criterion. In contrast to the excitatory masking of Fig. 1A, Figure 1B illustrates suppressive masking. The masker produces no increment in discharge rate over spontaneous, but shifts the rate-level function for the signal towards high intensities, resulting in a threshold elevation. Excitatory masking and suppressive masking are not mutually exclusive, as shown in Fig. 1C. The masker is excitatory, and suppresses the response to the signal because the rate for the signal plus masker is lower than the rate for the signal alone over a range of signal levels.

One can discriminate between the three forms of masking shown in Fig. 1A-C by measuring three types of thresholds from auditory-nerve fibers for each masker and signal frequency. Two of these thresholds have already been defined: These are the threshold in quiet, and the simultaneous masked threshold. The third threshold is the signal level for which the discharge rate
Physiological mechanisms of masking

in response to the signal alone exceeds the response to the masker alone by criterion. This "nonsimultaneous" threshold is similar in concept to the pulsation threshold in psychophysics (Houtgast, 1974). Like the pulsation threshold, it is not, strictly speaking, a masked threshold because the signal is detectable below threshold. The key point is that the difference between the simultaneous threshold and the nonsimultaneous threshold gives a measure of the contribution of suppression to masking because there is no suppression when the signal and the masker are not simultaneously present. Figure 1A shows that, when masking is excitatory, the simultaneous threshold is slightly lower than the nonsimultaneous threshold, and both masked thresholds exceed the threshold in quiet. In contrast, when masking is suppressive (Fig. 1B), the simultaneous threshold is higher than the nonsimultaneous threshold, and the threshold in quiet coincides with the nonsimultaneous threshold. When masking is both excitatory and suppressive (Fig. 1C), the simultaneous threshold is above the nonsimultaneous threshold, and both masked thresholds exceed the threshold in quiet

Method

Methods for recording from auditory-nerve fibers in anesthetized cats were basically as described by Kiang et al. (1965). For each fiber, three threshold measurements were made for a range of signal frequencies centered at the CF. The masker was always a 1-kHz tone at either 60 or 80 dB SPL. Figure 2A shows the method for measuring simultaneous masked thresholds, which mimics the two-tone, two-alternative forced choice paradigm of psychophysics. A pair of stimuli is presented repeatedly in random order. One stimulus is a 50-ms burst of sound consisting of a fixed 1-kHz masker and a variable tone signal, while the other stimulus is the masker alone. The signal level is adjusted by means of a PEST procedure (Taylor and Creelman, 1967) so that the spike count in response to the signal plus the masker exceeds the count for the masker alone for 75% of the presentations. Figure 2B shows the method for measuring nonsimultaneous masked thresholds. The stimulus pair now consists of the masker alone and the signal alone, and the signal level is adjusted by PEST so that the spike count for the signal exceeds the count for the masker 75% of the time. Thresholds in quiet were measured by the same method as either masked threshold except that the masker was omitted.

Results

Figure 3A shows the threshold in quiet, the simultaneous masked threshold, and the nonsimultaneous threshold as a function of signal frequency for three auditory-nerve fibers from one cat. The masker was a 1-kHz tone at 60 dB

2. More precisely, because the difference between simultaneous and nonsimultaneous thresholds is slightly negative for excitatory masking, this difference might remain negative even when there is a small suppression. However, a positive difference always indicates a contribution of suppression to masking.
Figure 2. Stimulus paradigms used measuring simultaneous and nonsimultaneous masked thresholds of auditory-nerve fibers.

Figure 3. Threshold in quiet, for simultaneous masked threshold, and nonsimultaneous threshold as a function of signal frequency for three auditory-nerve fibers from one cat, and for two intensities of a 1-kHz masker.

SPL. The CF of the center fiber is 1.2 kHz, close to the masker frequency. For this fiber, thresholds in both masking conditions are elevated by about 30 dB over thresholds in quiet. Masking is much smaller for the other two fibers, whose CF's are far from the 1-kHz masker frequency. For both the 0.5-kHz and the 1.2-kHz fibers, masking is largely excitatory because thresholds in the simultaneous and nonsimultaneous conditions are similar. In contrast, masking is suppressive for the 4.6-kHz fiber, because the simultaneous masked thresholds exceed thresholds in quiet by about 10 dB near the CF, while the nonsimultaneous thresholds nearly coincide with thresholds in quiet.

Figure 3B shows masked thresholds for the same fibers as in Fig. 3A when the level of the 1-kHz masker is raised to 80 dB SPL. For the 1.2-kHz fiber, simultaneous masked thresholds were greater than 95 dB SPL, except for 3 signal frequencies near the CF³. For the 0.5-kHz fiber, the 20-dB increase in masker level results in only a 12-dB increase in simultaneous masked

³ This fiber was "lost" before nonsimultaneous thresholds could be measured.
Physiological mechanisms of masking

thresholds, and even less for nonsimultaneous thresholds. In contrast, for the 4.6-kHz fiber, there is a 40-dB increase in simultaneous masked thresholds for signal frequencies near the CF. Nonsimultaneous thresholds are considerably lower than simultaneous thresholds, suggesting that masking is primarily suppressive for this fiber.

In order to relate these physiological data to psychophysical masking patterns, we need to examine masked thresholds of auditory-nerve fibers as a function of their points of innervation along the cochlea. It seems likely that, for any given signal and masker, it is the fibers with the lowest masked thresholds that provide most information for detecting the signal. Therefore, for each masker level and signal frequency, a V-shaped curve was fit to the patterns of fiber thresholds against CF. The tip of the V defines the "best threshold" for that signal frequency. Figure 4 shows best thresholds as a function of signal frequency, in quiet, and for both a simultaneous and a nonsimultaneous 1-kHz masker at 60 dB SPL. Best thresholds in simultaneous masking are maximum for signal frequencies near the 1-kHz masker, and decay more gradually from maximum for signal frequencies above the masker frequency than for frequencies below the masker, as shown by Sinex and Havey (1986). Best thresholds in the nonsimultaneous condition are below simultaneous thresholds over a broad range of signal frequencies both below and above 1 kHz. Thus, suppression contributes to masking for these frequencies. However, for signal frequencies near the masker, nonsimultaneous thresholds exceed simultaneous threshold, indicating that masking is excitatory in that range. Overall, the nonsimultaneous masking pattern is more sharply tuned than the simultaneous pattern.

Figure 4. Best thresholds of auditory-nerve fibers as a function of signal frequency, in quiet, and for both simultaneous and nonsimultaneous 1-kHz masker at 60 dB SPL. Each data point is obtained from threshold measurements in at least 10 auditory-nerve fibers.

Figure 5. Same as Fig. 4 for an 80-dB masker. The available data did not allow reliable estimates of best nonsimultaneous thresholds for signal frequencies near 1 kHz.
Figure 5 shows best thresholds patterns when the level of the 1-kHz masker is raised to 80 dB. At this level, simultaneous masking extends much farther toward the high frequencies than at the lower level. For signal frequencies near and below 1 kHz, simultaneous masking grows by 15 to 20 dB for the 20 dB increase in masker level. In contrast, for signal frequencies well above 1 kHz, the difference in thresholds between the two levels reaches 40 to 50 dB. This supralinear growth of physiological masking for signal frequencies above the masker has been observed by Sinex and Havey (1986). A key question is whether it is due to an increase in suppression or an increase in excitation. Figure 5 shows that, for the 80 dB masker, nonsimultaneous thresholds are about 40 dB below simultaneous thresholds for signal frequencies well above the masker. This threshold difference, which we attribute to suppression, is only 5 to 10 dB for the 60-dB masker (Fig. 4). Therefore, it seems that the supralinear growth of masking is primarily due to the rapid growth of suppression rather than to the growth of excitation. This result fits well with the fact that the rate of growth of two-tone rate suppression with suppressor level is about 2 dB/dB for suppressors below the CF (Abbas and Sachs, 1976; Delgutte, 1986; Costalupes et al., 1987).

Discussion

The goal of this study was to identify the contributions of suppression and spread of excitation to tone-on-tone masking by comparing the masked thresholds of auditory-nerve fibers measured with simultaneous and nonsimultaneous techniques. For 1-kHz maskers at 60 and 80 dB SPL, simultaneous masked thresholds were above nonsimultaneous thresholds over most of the range of signal frequencies, with the possible exception of the immediate vicinity of the masker frequency (Fig. 4 and 5). This shows that physiological masking is, in general, both excitatory and suppressive. Excitatory masking dominates for signal frequencies near and below the masker frequency, although suppression also contributes somewhat below the masker frequency. Suppressive masking dominates for signal frequencies well above the masker, particularly with the 80-dB masker.

The role of suppression in masking is complicated by the possibility that an off-CF signal might suppress the response to a simultaneous masker, as illustrated in Fig. 1D. The masker produces an increment in rate over spontaneous. When the signal level is raised in the presence of the masker, discharge rate first decreases due to suppression by the signal, then increases when the signal becomes excitatory. The masked threshold is now the level at which the rate for the signal plus masker is lower than the rate for the masker by criterion. The masked threshold is below the threshold in quiet, indicating that signals which, by themselves, cannot be detected, become detectable when the masker is introduced. This "unmasking" phenomenon should be taken into account in estimating best thresholds. For about 20 auditory-nerve fibers, we measured "unmasking thresholds" by substituting a probability
Physiological mechanisms of masking

criterion of 25% for that of 75% in the PEST procedure, and reversing the direction of level changes. With a 1-kHz simultaneous masker at 60 dB SPL, unmasking thresholds were above the best threshold curve of Fig. 4 for all fibers and all signal frequencies. On the basis of these preliminary data, it does not appear likely that taking into account the suppression of the masker response by the signal would greatly change the physiological masking patterns of Fig. 4 and 5. A similar conclusion was reached by Sinex and Havey (1986).

Our conclusions concerning the role of suppression in simultaneous masking differ somewhat from those of Pickles (1984). By comparing the discharge rates of auditory-nerve fibers at masked threshold with the rate at the threshold in quiet, Pickles concluded that masking is excitatory, except for a narrow range of signal frequencies below the masker frequency in which masking is suppressive. However, a greater rate at masked threshold than at the threshold in quiet only implies that masking is partly excitatory, and does not rule out a suppressive component (Fig. 1C). Thus Pickles’ results are consistent with our conclusion that masking is generally both excitatory and suppressive.

The physiological masking patterns of Fig. 4 and 5 resemble psychophysical masking patterns in many respects (Wegel and Lane, 1924; Houtgast, 1974). Both have a maximum near the masker frequency, and a pronounced skew towards high frequencies. Both physiological and psychophysical masking patterns are more sharply tuned in nonsimultaneous than in simultaneous masking4. In simultaneous masking, both physiological and psychophysical masked thresholds grow faster than linearly with masker level for signal frequencies above the masker, a phenomenon called “upward spread of masking”. Because nonsimultaneous thresholds (which do not include effects of suppression) do not grow so rapidly with masker level, we concluded that the upward spread of simultaneous masking is due primarily to the supralinear growth of suppression rather than to the growth of excitation. If true, this conclusion has implications for the interpretation of psychophysical experiments. For example, Zwicker (1970) explained the decrease in pure-tone intensity difference limens (DL) with increasing stimulus level in terms of the nonlinear growth of simultaneous masking patterns. Our results suggest that this interpretation is incorrect because the nonlinear growth in masking is due to suppression, which does not occur for single tones. Florentine and Buus (1981) have provided an alternative explanation for the decrease in DL with intensity which does not require a nonlinear growth in excitation patterns.

In conclusion, both suppression and spread of excitation are important for explaining psychophysical data on simultaneous masking in terms of physiological mechanisms. Specifically, our results suggest that spread of

4. Lufti’s (1988) recent psychophysical data show roughly the same sharpness of tuning in simultaneous and in forward masking. However, his results are limited to signal frequencies within a half-octave of the masker frequency, whereas we find the largest difference between the two masking conditions for signal frequencies well above the masker.
excitation determines the overall shape of masking patterns, while suppression is largely responsible for the upward spread of simultaneous masking and for differences between simultaneous and nonsimultaneous masking. This implies that psychophysical masking patterns obtained by simultaneous techniques can only provide a crude representation of the pattern of activity produced by the masker at peripheral stages of the auditory system because these masking patterns also reflect suppression of the signal by the masker. This representation will be particularly distorted when the masker includes intense low-frequency components such as the first formant of speech.

Acknowledgement
This research was supported by NIH grant NS13126. I thank L.D. Braida, M.C. Brown, J.L. Goldstein, and N.Y.S. Kiang for their valuable comments.

References

Comments

Moore:

At first sight, there appears to be a discrepancy between your results and the view commonly held by psychophysicists that the effects of suppression are not revealed in simultaneous masking (Houtgas, 1974; for a review see Moore and O'Loughlin, 1986); you suggest that suppression is largely responsible for the upward spread of simultaneous masking. I think that the discrepancy can be resolved if we postulate that the excitation pattern for a pure tone (the masker in this case) can be sharpened by suppression; activity at the peak of the pattern suppresses weaker activity on the skirts, an effect which we can call self-suppression. Your masking patterns determined in 'nonsimultaneous masking' include the effects of this self-suppression. Indeed, when only a single pure tone is presented at a time, physiological estimates of frequency selectivity always include the effects of self-suppression. In simultaneous masking the suppression is applied both to the signal and to the masker. This does not change the signal-to-masker ratio in any channel (relative to what it would be without suppression), but the overall effect is that the signal threshold is higher in simultaneous than in nonsimultaneous masking. In summary, I would interpret your results in this way: Masking patterns are sharply tuned in nonsimultaneous masking since the effects of self-suppression are revealed. Masking patterns in simultaneous masking are less sharply tuned since the suppression affects both the masker and the signal; the simultaneous masking patterns can be thought of as revealing what the excitation pattern of the masker would be like if no suppression were occurring.